FEBS Letters 584 (2010) 1205–1211

FEBS Letters

journal homepage: www.FEBSLetters.org

Recognition of cellooligosaccharides by a family 28 carbohydrate-binding module

Kazuhide Tsukimoto [a,1], Rie Takada [b,1], Yuko Araki [c], Kentaro Suzuki [a], Shuichi Karita [d], Takayoshi Wakagi [a], Hirofumi Shoun [a], Takashi Watanabe [b], Shinya Fushinobu [a,*]

[a] Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[b] Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
[c] Graduate School of Bioresources, Mie University, 1577 Kurimamachiya, Tsu 514-8507, Japan
[d] Graduate School of Regional Innovation Studies, Mie University, 1577 Kurimamachiya, Tsu 514-8507, Japan

ARTICLE INFO

Article history:
Received 7 January 2010
Revised 8 February 2010
Accepted 10 February 2010
Available online 14 February 2010

Edited by Hans Eklund

Keywords:
Crystal structure
Carbohydrate-binding module family 28
Cellooligosaccharide
Endoglucanase
Cellulase
Clostridium josui

ABSTRACT

The crystal structures of a carbohydrate-binding module (CBM) family 28 domain of endoglucanase Cel5A from *Clostridium josui* have been determined in ligand-free and complex forms with cellobiose, cellotetraose, and cellopentaose as the first complex structures of this family. In the cleft of a β-sandwich fold, the ligands are recognized by stacking interactions and hydrogen bonds. Conformations of the bound cellooligosaccharides are similar to those in crystals and solution but clearly different from the cellulose structure. Interestingly, the glucan chain bound on CBM28 is in the opposite direction of that bound to CBM17, although these families share significant structural similarity.

1. Introduction

Carbohydrate-binding modules (CBMs) are non-catalytic modules that bind to poly- or oligosaccharides. One or more CBMs are often attached to a catalytic module of plant cell wall polysaccharide-degrading enzymes. To date, more than 50 CBM families have been identified [1], and their binding modes have been classified into three types [2]: type A CBMs have a flat hydrophobic surface that interacts with crystalline surfaces; type B CBMs have a cleft to accommodate single glycan chains, and type C CBMs have pockets to accommodate oligosaccharides.

The endoglucanase Cel5A from *Clostridium josui* consists of 930 amino acids and contains a catalytic glycoside hydrolase (GH) family 5 module, a CBM family 17 (CBM17 from *Clostridium josui* Cel5A, *Cj*CBM17), a CBM family 28 (CBM28 from *Clostridium josui* Cel5A, *Cj*CBM28), and three repeats of a surface-layer homology module (GenBank Accession# BAA12826). *C. josui* produces a cellulosome, a cellulase complex, and surface-layer homology module-containing enzymes that are displayed on its cell surface [3]. Cel5A is among the most commonly exposed cellulolytic enzymes and is more active on non-crystalline cellulose than crystalline one [4,5]. Cel5A can hydrolyze cellooligosaccharides, carboxymethyl cellulose, and ball-milled cellulose, but cannot hydrolyze cellobiose, cellotriose, or xylan. A microcrystalline cellulose, Avicel is hydrolyzed slightly. Binding of individual domains of *Cj*CBM17, *Cj*CBM28 and tandem domains of *Cj*CBM17/28 to various ligands has been characterized [6]. Both of the individual CBMs have at least four binding subsites since cellotriose binding was not observed. The binding constants (K_a) for cellooligosaccharides are similar: *Cj*CBM17, K_a(cellotetraose) = $0.2 \times 10^5\ M^{-1}$ and K_a(cellopentaose) = $5.4 \times 10^5\ M^{-1}$; and *Cj*CBM28, K_a(cellotetraose) = $0.7 \times 10^5\ M^{-1}$ and K_a(cellopentaose) = $5.2 \times 10^5\ M^{-1}$. However, their binding thermodynamics are different; the binding of *Cj*CBM28 is fully driven by large negative enthalpy, while *Cj*CBM17 is partially driven by favorable changes in entropy.

CBM17 and CBM28 share slight sequence similarities and both have β-sandwich folds with significant structural similarities [7]. These families belong to type B CBM and have a shallow groove that binds the cellooligosaccharide with two conserved Trp residues. However, other amino acid residues involved in ligand

Abbreviations: CBM, carbohydrate-binding module; GH, glycoside hydrolase; *Cj*CBM17, CBM17 from *Clostridium josui* Cel5A; *Cj*CBM28, CBM28 from *Clostridium josui* Cel5A; *Cc*CBM17, CBM17 from *Clostridium cellulovorans* EngF; *Bsp*CBM28, CBM28 from *Bacillus* sp. 1139 GH5 endoglucanase; RMSD, root mean square deviation

* Corresponding author. Fax: +81 3 5841 5151.
E-mail address: asfushi@mail.ecc.u-tokyo.ac.jp (S. Fushinobu).
[1] These authors contributed equally to this work.

0014-5793/$36.00
doi:10.1016/j.febslet.2010.02.027

binding are not conserved. The crystal structures were previously solved for CBM17 from *Clostridium cellulovorans* EngF (*Cc*CBM17; 61% identity to *Cj*CBM17), in complex with cellotetraose [8], and CBM28 from *Bacillus* sp. 1139 GH5 endoglucanase Cel5A (*Bsp*CBM28; 48% identity to *Cj*CBM28), in a ligand-free form [7]. Recently, fluorescently-labeled *Cj*CBM28 was used to analyze the complex surfaces of pretreated wood biomass, revealing that *Cj*CBM28 is site-specifically absorbed on external fibrous structures [9]. However, interactions of CBM28 with a glycan chain are still unclear, because a ligand-bound crystal structure has not been available. In this report, we present the crystal structures of *Cj*CBM28 complexed with cellooligosaccharides as the first three-dimensional view of the ligand interactions of CBM28.

2. Materials and methods

2.1. Protein preparation and crystallography

N-Terminally His-tagged *Cj*CBM28 protein (MRGSHHHHHHR and amino acid number 561–752 of Cel5A) was expressed using pQE30 vector (Qiagen) and *Escherichia coli* M15, and then purified using Ni-NTA agarose (Qiagen) according to the manufacturer's instructions. Crystals were prepared by sitting-drop vapor diffusion method at 4 °C, by mixing 1 μL of 20 mg/mL protein in 50 mM Tris–HCl (pH 7.0) and 1 μL of reservoir solution. Ligand-free crystals were grown with reservoir solution consisting of 25% PEG MME 2000, 0.2 M $(NH_4)_2SO_4$, and 0.2 M Na–acetate (pH 4.6). Co-crystals with cellooligosaccharides were grown in the presence of 5 mM cellotriose, 0.5 mM cellotetraose, or 0.5 mM cellopentaose with reservoir solution consisting of 20% PEG 8000 and 50 mM KH_2PO_4. After cryoprotection with 20% glycerol, the crystals were flash-cooled in a nitrogen stream at 100 K. Diffraction data were collected using beamlines at the Photon Factory, KEK, Tsukuba (λ = 1.0 Å). Diffraction images were processed using HKL2000 [10]. Molecular replacement was performed with MOLREP [11], using the *Bsp*CBM28 structure as a search model. Model rebuilding and refinement were performed using Coot [12] and Refmac5 [13]. Data collection and refinement statistics are shown in Table 1. Figures were prepared using PyMol (DeLano Scientific).

3. Results and discussion

3.1. Overall structure

Ligand-free and cellooligosaccharide complex structures were determined at 1.4–1.6 Å resolutions. All crystals belong to the same space group, but the cell constants and growth conditions were clearly different between ligand-free and complex crystals (Table 1). The asymmetric unit of all structures contained one *Cj*CBM28 molecule (Arg11–Glu203 starting from the last residue in the tag sequence to the C-terminus of the construct) and one Ca^{2+} ion. The structures contained a few solvent molecules (sulfate, phosphate, and glycerol), but they did not influence ligand recognition. *Cj*CBM28 has a typical β-sandwich fold, and the cellooligosaccharide ligand binds in a cleft formed by this fold (Fig. 1A). A metal ion bound at the back of the β-sandwich fold (green sphere) was observed, although buffers used during the protein purification steps contained no divalent cations. This position is conserved with the Ca^{2+}-binding site of *Bsp*CBM28. The refined temperature factor for the Ca^{2+} ion was normal (8.8–18.8 $Å^2$ in all structures), and the difference electron density was almost flat. When a Ni^{2+} ion was placed in this site, a large negative peak appeared in the difference map (data not shown). The Ca^{2+} ion is coordinated by the main chain carbonyl groups of Thr31 and Lys61, Oε2 of Glu33, Oδ1 and Oδ2 of Asp198, Oγ of Ser60, and a water molecule.

The structure of *Cj*CBM28 is very similar to *Bsp*CBM28 (Fig. 1B). The root mean square deviation (RMSD) between the ligand-free structures is 0.67 Å for 161 Cα atoms. However, the visible N-terminal region of the *Cj*CBM28 structure is longer than that of *Bsp*CBM28 by 7 amino acid residues. In addition, the following regions exhibit significant Cα deviations (>2 Å) on superposition; Gly87–Asn90, Ser144–Gly152, and Asp163–Leu171. The first and third regions are located at the peripheries of the cleft (Fig. 1B, asterisks), but they do not contribute to ligand recognition. Both

Table 1
Data collection and refinement statistics.

Data set	Ligand-free	Cellobiose	Cellotetraose	Cellopentaose
A. Data collection				
PDB code	3ACF	3ACG	3ACH	3ACI
Beamline	BL5A	NE3A	BL5A	BL5A
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	*a* = 47.164	*a* = 39.468	*a* = 39.576	*a* = 39.568
	b = 63.305	*b* = 63.820	*b* = 63.900	*b* = 63.879
	c = 71.251	*c* = 75.449	*c* = 75.619	*c* = 75.617
Resolution (Å)	50.00–1.60 (1.63–1.60)	50.00–1.50 (1.53–1.50)	50.00–1.40 (1.42–1.40)	50.00–1.60 (1.63–1.60)
Total reflections	189 654	216 155	223 207	182 203
Unique reflections	28 716	31 189	35 569	25 798
Completeness (%)	99.6 (99.9)	99.9 (99.7)	92.3 (95.5)	99.2 (97.1)
Redundancy	6.6 (6.8)	6.9 (5.2)	6.3 (6.4)	7.1 (7.2)
Mean I/σI	34.4 (6.7)	25.5 (2.7)	38.7 (5.2)	42.5 (11.1)
R_{sym} (%)	6.6 (21.7)	8.0 (37.6)	4.7 (24.7)	5.3 (14.8)
B. Refinement				
Resolution (Å)	28.93–1.60	27.27–1.50	27.34–1.40	31.94–1.60
No. of reflections	27 198	29 563	33 734	24 444
R-factor/R_{free} (%)	15.6/19.1	14.7/19.0	16.2/19.8	14.1/18.8
No. of atoms	1819	1944	1898	1933
No. of solvents	1 (SO_4)	2 (PO_4) and 1 (glycerol)	1 (PO_4)	2 (PO_4)
RMSD from ideal values				
Bond lengths (Å)	0.030	0.030	0.028	0.027
Bond angles (°)	2.408	2.507	2.354	2.278
Ramachandran plot (%)				
Favored	97.4	97.9	97.9	97.9
Allowed	2.6	2.1	2.1	2.1
Outlier	0.0	0.0	0.0	0.0

proteins have three aromatic residues in the cleft (Trp78, Trp129, and Phe128 in *Cj*CBM28; and Trp68, Trp119, and Tyr118 in *Bsp*CBM28), forming a long hydrophobic patch. At the position of Gly87 in *Cj*CBM28, *Bsp*CBM28 has an additional aromatic residue (Trp77). As suggested by Jamal et al. [7], *Bsp*CBM28 possibly binds cellooligosaccharide ligands in an alternative mode that involves the stacking interaction with Trp77 (Supplementary Fig. S1). This may explain the slightly different thermodynamic characteristics of ligand binding between *Cj*CBM28 and *Bsp*CBM28 [14].

3.2. Cellooligosaccharide recognition

The ligand-free and the three complex structures are almost identical, and there are no significant displacements even for the ligand-interacting side chains in the cleft (Supplementary Fig. S2). RMSDs between the ligand-free and complex structures (0.39–0.41 Å for Cα and 0.57–0.71 Å for all atoms) are slightly higher than those between the complex structures (0.08–0.15 Å for Cα and 0.28–0.41 Å for all atoms), but this discrepancy is mainly due to the change in crystal packing. These findings suggest that no conformational change from its native ligand-free state is required for CBM28 to bind cellooligosaccharide ligands.

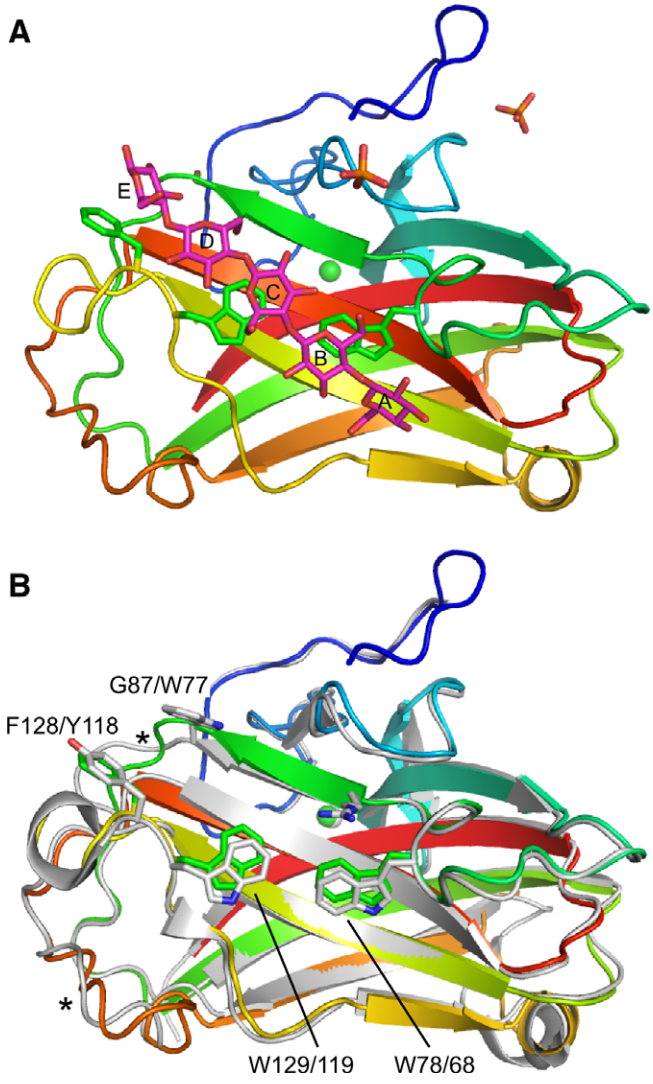

Fig. 1. (A) Rainbow-colored overall structure of the *Cj*CBM28–cellopentaose complex and (B) superimposition with *Bsp*CBM28 (grey). The Ca^{2+} and phosphate ions are shown as sphere and stick models, respectively. The cellopentaose and aromatic residues in the cleft are shown as stick models. (B) Overlapping residues for the two structures are labeled as *Cj*CBM28/*Bsp*CBM28. Regions consisting of Gly87–Asn90 and Asp163–171 are indicated by asterisks.

The electron density maps for cellooligosaccharides were clearly observed (Fig. 2). However, density peaks for only two glucose units were observed from the dataset for the crystals grown in the presence of cellotriose (Fig. 2A). We concluded that the crystals contained cellobiose, which is contaminant in reagent or degradation products, since no electron density peaks for a partially disordered glucose unit at either flanking end were observed. We designated the subsites for cellopentaose as A–E from the non-reducing end (Fig. 1A). Cellotetraose and cellobiose occupy subsites B–E and B–C, respectively, revealing different affinities of each subsite for a glucose unit. Subsites B, C, and E form stacking interactions with aromatic residues (Trp78, Trp129, and Phe128), and these subsites likely have higher affinities than the others. These observations are consistent with our previous report, which showed that cellotriose binding was not observed by isothermal titration calorimetry measurements [6]. In this study, high concentrations of cellotetraose/cellobiose present in the crystallization drop or the specific crystallization conditions might have promoted crystallization of cellobiose-bound subpopulations of the protein. The growth conditions for ligand-free and complex crystals were clearly different, and no crystals grew under the latter condition without cellooligosaccharides. The stacking interactions present between the aromatic residues and the pyranose rings are consistent with the observation of a large UV absorbance change for *Bsp*CBM28 on cellooligosaccharide binding, which indicates involvement of Trp side chains in that process [15].

Schematic drawing of the interaction of *Cj*CBM28 with cellopentaose is shown in Supplementary Fig. S3. In the cellopentaose complex, the side chains of Asp76, Arg83, Gln131, Asp135, and Arg178, and the main chain atoms of Gly77 and Gly127, form direct hydrogen bonds with hydroxyl groups on both sides of cellooligosaccharides. These hydrophilic residues are highly conserved in CBM28. Recognition at subsite A is responsible to the increase in the binding affinity to cellopentaose than cellotetraose ($\Delta\Delta G$ = −1.2 kcal mol^{-1}) [6]. The two direct hydrogen bonds formed by Asp135 and Gly77 appear to contribute to $\Delta\Delta H$ of −12.5 kcal mol^{-1}, but the negative $T\Delta\Delta S^0$ value (−11.3 kcal mol^{-1}) may be partly due to the immobilization of a larger number of conformers adopted by the longer ligand. The hydroxyl groups of cellooligosaccharides bind numerous water molecules (red spheres in Fig. 2). However, in the ligand-free form, the side chains of hydrophilic residues bind only a few waters, and no ordered water molecules were observed on the aromatic side chains. These observations explain the unfavorable entropy change on ligand binding [6], because an excess number of waters seem to be constrained on ligand binding.

3.3. Comparison with CBM17

Fig. 3 shows a comparison between the *Cj*CBM28–cellopentaose and *Cc*CBM17–cellotetraose complex [8] structures. Both CBM families have the ligand binding site on the concave surface formed by five strands of the β-sandwich fold. The subsites of *Cc*CBM17 (Fig. 3B, designated A–D from the non-reducing end) do not overlap with those of *Cj*CBM28 (Fig. 3A). A conserved Trp residue forms subsite C in both structures (Trp129 in *Cj*CBM28 and Trp135 in *Cc*CBM17), but they do not overlap due to the different main chain and side chain conformations. The conserved Trp88 residue in *Cc*CBM17 is located in a significantly different position from the corresponding Trp78 in *Cj*CBM28 because of the presence of a short α-helix (blue in Fig. 3B), making its groove shallower (Fig. 3D). Cellotetraose bound in *Cc*CBM17 is also recognized by direct hydrogen bonds with hydrophilic residues (Fig. 3B), which are

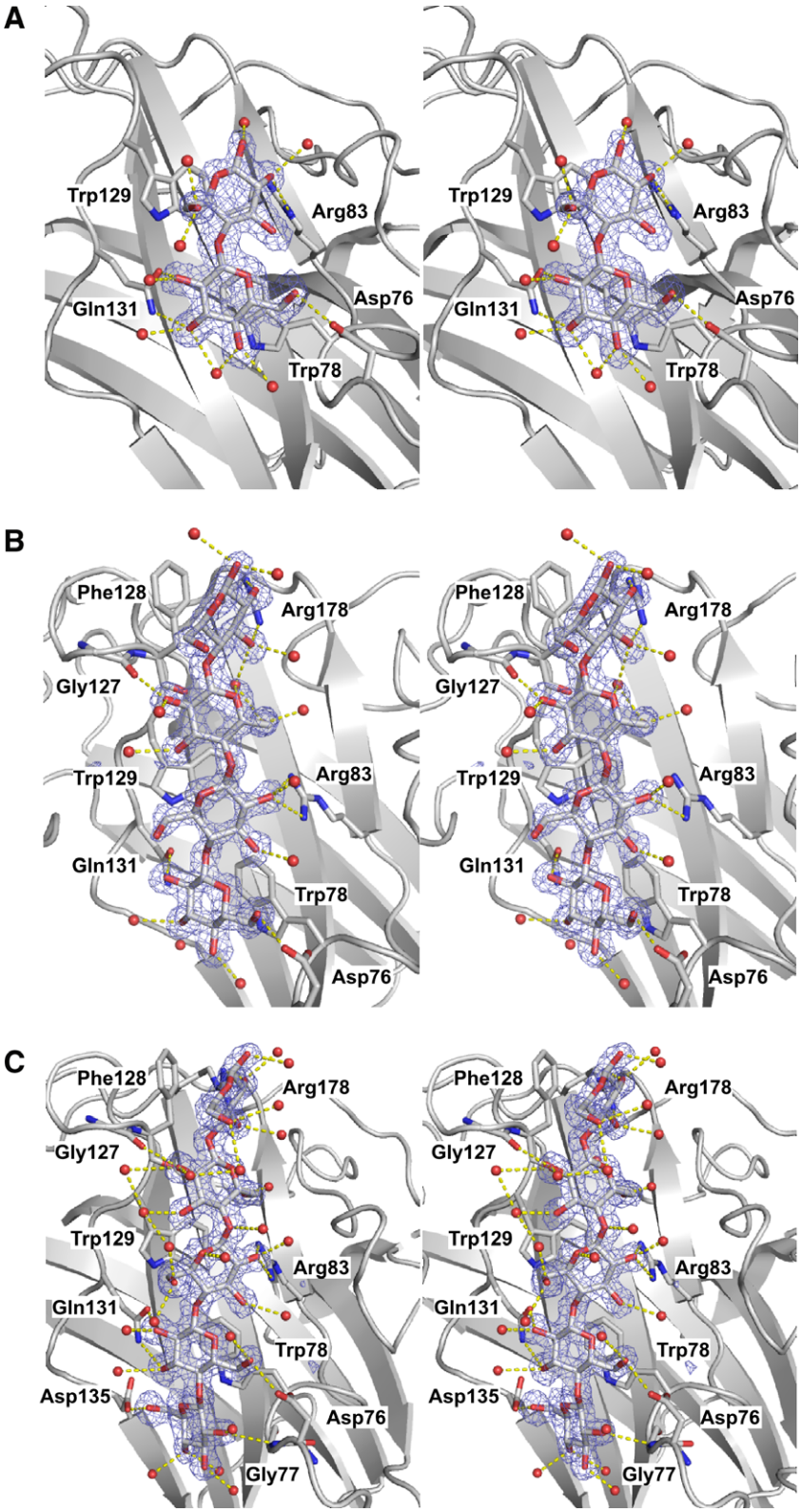

Fig. 2. Stereoviews of *Cj*CBM28 interactions with cellobiose (A), cellotetraose (B), and cellopentaose (C). Omit $|F_o| - |F_c|$ maps (3.0σ) are shown.

conserved within the CBM17 family. *Cc*CBM17 exhibits the optimal affinity for cellohexaose than cellopentaose or cellotetrase, indicating that it has six subsites [8]. Notenboom et al. suggested presence of the fifth subsite at the reducing-end side of the cellotetraose near Trp88. Interestingly, the glucan chains interacting with these CBMs run in opposite directions. Because both CBM families specifically recognize the hydroxyl groups of cellooligosaccharides, they must have directional specificity for glucan chain binding. To date, most CBM28 domains are found downstream of CBM17 domains. The two type B CBMs with different chain

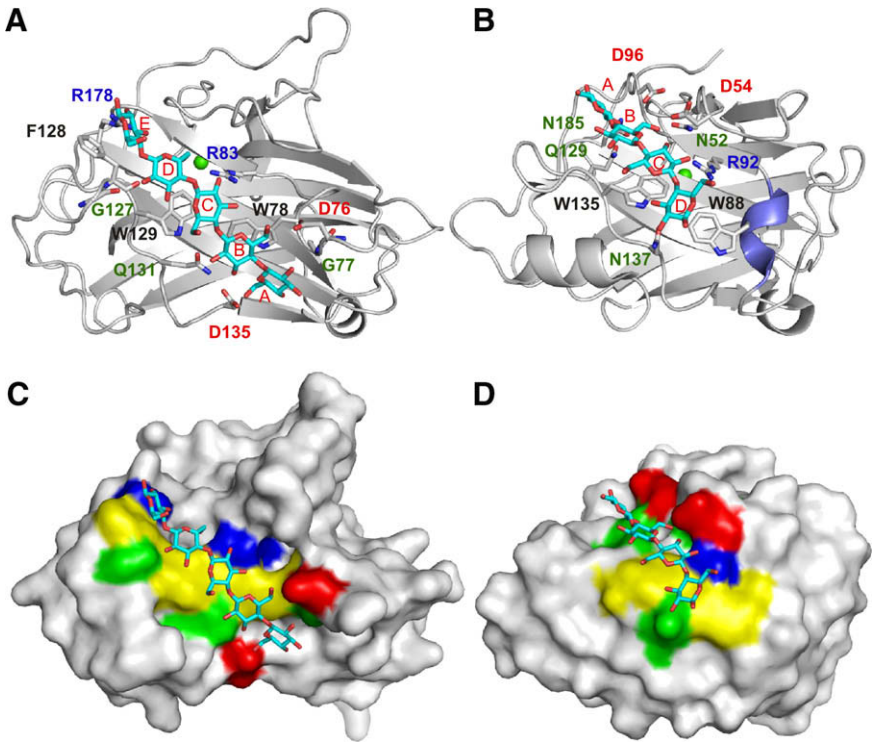

Fig. 3. Comparison of *Cj*CBM28–cellopentaose complex (A and C) with *Cc*CBM17–cellotetraose complex (B and D). (A and B) Ribbon representation with important residues for ligand binding shown as a stick model. The subsites are labeled alphabetically from the non-reducing end. An α-helix, in which Trp88 of *Cc*CBM17 is located, is colored blue. (C and D) Molecular surfaces. Important residues in the cleft are colored yellow (aromatic), red (acidic), blue (basic), and green (neutral).

directional and conformational specificities may work effectively in tandem on natural cellulosic substrates, as they bind to different regions of non-crystalline cellulose [14].

The binding mode of CBM4 is significantly different from those of CBM28 and CBM17 [16], although it is also a type B CBM having a β-sandwich fold. In the cleft of CBM4, a glucan chain is bound in an edge-on mode, in which the hydroxyl groups on one side point into the core of the protein. Therefore, the glucan chain on CBM4 is rotated ~90° along its long axis relative to those on CBM28 and CBM17. Interestingly, CBM4 domains bind cellooligosaccharide derivatives in multiple orientations, indicating that there is no directional preference for ligand binding [17].

3.4. Conformations of the bound ligands and the binding site architectures

Table 2 summarizes the conformations of cellooligosaccharides bound to *Cj*CBM28 and *Cc*CBM17. All sugar rings are in the $^4C_1$ chair conformation. Most of the conformations of the hydroxymethyl groups in *Cj*CBM28 are *gg* except for subsite B, in which the hydrogen bond with Asp76 fixes the conformation to *gt*. Therefore, the cellooligosaccharides do not form inter-sugar hydrogen bond between O2′–O6, in contrast to the cellulose Iβ structure whose sugars are all in the *tg* conformation [18]. In *Cc*CBM17, all of the hydroxymethyl groups of cellotetraose are modeled as *gt*, but that in subsite B exclusively forms hydrogen bonds with the protein [8]. Cellobiose and its analogs mostly adopt the *gg* or *gt* conformations in solution or crystals [19,20].

The glycosidic torsion angles of the cellooligosaccharides are within the lowest-energy region on the energy surface map of cellobiose ($-100° < \varphi < -65°$ and $-150° < \psi < -115°$) [21]. However, the torsion angles are aggregated in two separate regions on the φ–ψ plot (Supplementary Fig. S4). Angles between sugar ring planes are about 180° for glycosidic bonds between subsites C and D in *Cj*CBM28 and that between subsites B and C in *Cc*CBM17. The glycosidic torsion angles at these bonds are similar to those of cellulose Iβ and II ($\varphi = -88° \sim -99°$ and $\psi = -142° \sim -150°$) [18,22], in which a twofold screw axis symmetry exists. Angles between the sugar ring planes connected by other glycosidic bonds are 130–150° and thus fit on the non-flat platform formed by the aromatic residues in the cleft. In *Cj*CBM28, the angles formed between Trp78 and Trp129, and between Trp129 and Phe128, are about 120° and 140°, respectively. Therefore, glycosidic bonds of the bound cellooligosaccharides take a twisted conformation except for that between subsites C and D (Fig. 2). The hydrogen bond between Gly127 and O2 of the sugar in subsite D appears to fix this glycosidic bond conformation. The inter-sugar hydrogen bonds between O5′–O3 are present in all positions, but it appears weaker between the sugars in subsites D and E.

Helix parameters calculated for trisaccharides indicate that the cellooligosaccharides form slightly left-handed helix ($n < -2.10$), which is similar to most cellooligosaccharides or its analogs in crystals (Supplementary Table S1) [20]. In summary, the cellooligosaccharides bound to the two CBMs adopt conformations basically similar to those found in crystals and solution, but they are clearly different from the cellulose structure. This result is consistent with the analysis using fluorescently-labeled *Cj*CBM28 that revealed specific binding to exposed surfaces near the bent or distorted parts of the pulp fibers [9].

3.5. Domain architecture of Cel5A

Truncation and mutagenesis work on *Bacillus* sp. 1139 Cel5A has shown that the binding abilities of both CBM domains are indeed correlated with the hydrolytic activity against regenerated cellulose [14]. Fig. 4 shows a schematic illustration of domain organization of *C. josui* Cel5A. GH5 endoglucanase from *Bacillus* sp. KSM-635 (62% identity to *Cj*CBM28), *Cc*CBM17, and *Cj*CBM28 are

Table 2
Conformations of cellooligosaccharides bound to *Cj*CBM28 and *Cc*CBM17.

	O6 conformation		Glycosidic torsion angles[b] and inter-sugar hydrogen bond distances			
	Subsite	χ (°)[a]	Subsites	ϕ (°)	ψ (°)	O5′–O3 (Å)
CBM28-cellopentaose	A	−42 (*gg*)	A–B	−70	−139	2.7
	B	65 (*gt*)	B–C	−75	−128	2.8
	C	−57 (*gg*)	C–D	−94	−146	2.7
	D	−62 (*gg*)	D–E	−75	−119	3.1
	E	−71 (*gg*)				
CBM28-cellotetraose	B	68 (*gt*)	B–C	−71	−135	2.7
	C	−61 (*gg*)	C–D	−89	−149	2.7
	D	−64 (*gg*)	D–E	−70	−119	3.1
	E	−66 (*gg*)				
CBM28-cellobiose	B	69 (*gt*)	B–C	−73	−131	2.9
	C	−65 (*gg*)				
CBM17-cellotetroase	A	50 (*gt*)	A–B	−76	−118	3.1
	B	42 (*gt*)	B–C	−99	−131	2.9
	C	68 (*gt*)	C–D	−69	−127	2.7
	D	18 (*gt*)				

[a] Torsion angle χ = O5–C5–C6–O6.
[b] Torsion angle ϕ = O5′–C1′–O4–C4, ψ = C1′–O4–C4–C5.

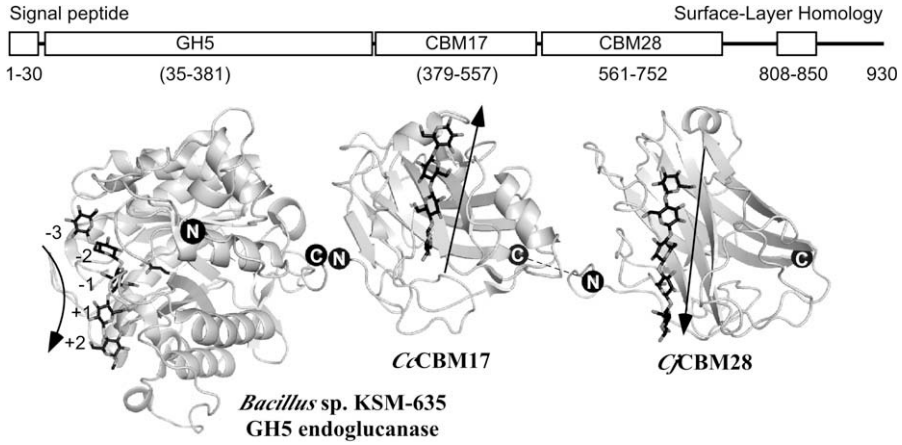

Fig. 4. Schematic illustration of domain organization of *C. josui* Cel5A. Structures of GH5 endoglucanase from *Bacillus* sp. KSM-635 (1G01) superimposed with methyl $4,4^{II},4^{III},4^{IV}$-tetrathio-α-cellopentoside bound to *Bacillus agaradhaerens* Cel5A (1H5 V), *Cc*CBM17–cellotetraose complex, and CBM28–cellopentaose are shown with their N- and C-termini (spheres). Numbers in parentheses are positions of Cel5A corresponding to the GH5 and CBM17 domain structures. Directions of the cellooligosaccharide ligands are shown by arrows from the non-reducing to reducing ends.

modeled by roughly connecting their N- and C-termini. The modeling illustrate that the glucan chains bound to these domains run in antiparallel orientation, and the chains can be separated about 30–50 Å. However, both of the CBM domains have a linker of at least 10 amino acid residues, and the relative orientations and/or positions of the domains can alter drastically due to conformational change of the linkers. The absorption constant K_a of tandem *Cj*CBM17/18 to ball-milled cellulose is approximately 3–7 times higher than that of individual domains, but the maximum amount of the bound CBM $[PC]_{max}$ decreases [6]. These data suggest the simultaneous interaction of the two domains and requirement of larger amorphous spaces for binding of *Cj*CBM17/18 than an individual CBM. The number of cellopentaose molecules bound to *Cj*CBM17/28 is 1.03, indicating the interference of each domain's binding activity [6]. Further investigation on the overall structure of this multi-domain enzyme is of interest.

Acknowledgements

We thank the staff of the Photon Factory for the X-ray data collection. This work was supported by The New Energy and Industrial Technology Development Organization.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2010.02.027.

References

[1] Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 37, D233–D238.
[2] Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769–781.
[3] Doi, R.H., Kosugi, A., Murashima, K., Tamaru, Y. and Han, S.O. (2003) Cellulosomes from mesophilic bacteria. J. Bacteriol. 185, 5907–5914.
[4] Fujino, T., Sukhumavasi, J., Sasaki, T., Ohmiya, K. and Shimizu, S. (1989) Purification and properties of an endo-1,4-β-glucanase from *Clostridium josui*. J. Bacteriol. 171, 4076–4079.
[5] Fujino, T., Sasaki, T., Ohmiya, K. and Shimizu, S. (1990) Purification and properties of an endo-1,4-β-glucanase translated from a *Clostridium josui* gene in *Escherichia coli*. Appl. Environ. Microbiol. 56, 1175–1178.
[6] Araki, Y., Karita, S., Tanaka, A., Kondo, M. and Goto, M. (2009) Characterization of family 17 and family 28 carbohydrate-binding modules from *Clostridium josui* Cel5A. Biosci. Biotechnol. Biochem. 73, 1028–1032.
[7] Jamal, S., Nurizzo, D., Boraston, A.B. and Davies, G.J. (2004) X-ray crystal structure of a non-crystalline cellulose-specific carbohydrate-binding module: CBM28. J. Mol. Biol. 339, 253–258.

[8] Notenboom, V., Boraston, A.B., Chiu, P., Freelove, A.C., Kilburn, D.G. and Rose, D.R. (2001) Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. J. Mol. Biol. 314, 797–806.
[9] Kawakubo, T. et al. (2010) Analysis of exposed cellulose surfaces in pretreated wood biomass usng carbohydrate-binding module (CBM)-cyan fluorescent protein (CFP). Biotechnol. Bioeng. 105, 499–508.
[10] Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326.
[11] Vagin, A. and Teplyakov, A. (2010) Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. 66, 22–25.
[12] Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (in press) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr.
[13] Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255.
[14] Boraston, A.B., Kwan, E., Chiu, P., Warren, R.A. and Kilburn, D.G. (2003) Recognition and hydrolysis of noncrystalline cellulose. J. Biol. Chem. 278, 6120–6127.
[15] Boraston, A.B., Ghaffari, M., Warren, R.A. and Kilburn, D.G. (2002) Identification and glucan-binding properties of a new carbohydrate-binding module family. Biochem. J. 361, 35–40.
[16] Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G. and Davies, G.J. (2002) Differential oligosaccharide recognition by evolutionarily-related β-1,4 and β-1,3 glucan-binding modules. J. Mol. Biol. 319, 1143–1156.
[17] Johnson, P.E., Brun, E., MacKenzie, L.F., Withers, S.G. and McIntosh, L.P. (1999) The cellulose-binding domains from *Cellulomonas fimi* β-1,4-glucanase CenC bind nitroxide spin-labeled cellooligosaccharides in multiple orientations. J. Mol. Biol. 287, 609–625.
[18] Nishiyama, Y., Langan, P. and Chanzy, H. (2002) Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. J. Am. Chem. Soc. 124, 9074–9082.
[19] Olsson, U., Serianni, A.S. and Stenutz, R. (2008) Conformational analysis of beta-glycosidic linkages in ^{13}C-labeled glucobiosides using inter-residue scalar coupling constants. J. Phys. Chem. B 112, 4447–4453.
[20] French, A.D. and Johnson, G.P. (2004) What crystals of small analogs are trying to tell us about cellulose structure. Cellulose 11, 5–22.
[21] French, A.D. and Johnson, G.P. (2004) Advanced conformational energy surfaces for cellobiose. Cellulose 11, 449–462.
[22] Langan, P., Nishiyama, Y. and Chanzy, H. (2001) X-ray structure of mercerized cellulose II at 1 Å resolution. Biomacromolecules 2, 410–416.